

Supporting Online Material for

A Genomewide Search for Ribozymes Reveals an HDV-like Sequence in the Human *CPEB3* Gene

Kourosh Salehi-Ashtiani, Andrej Lupták, Alexander Litovchick, Jack W. Szostak*

* To whom correspondence should be addressed. E-mail:
szostak@molbio.mgh.harvard.edu

Published 22 September 2006, *Science* **313**, 1788 (2006).
DOI: 10.1126/science.1129308

This PDF file includes

Materials and Methods
Figs. S1 to S9
References

Supporting Online Material.

Salehi-Ashtiani et al.

Materials and Methods.

Library construction.

Human genomic DNA (isolated from whole blood) was purchased from Clontech (Cat. 6550-1) and DNase I was from Boehringer Mannheim. The DNA was partially digested with DNase I and purified on agarose gels. The purified DNA was incubated with Taq DNA polymerase and dATP at 72 °C for 20 minutes to add non-templated 3'-A overhangs and subsequently ligated to 5'-phosphorylated hairpin oligonucleotides (SLLT and ALLT) with 3'-T overhangs using T4 DNA ligase at 18 °C for 1 hour (Fig. S1A). The ligated DNA was incubated with mung bean nuclease to remove the single-stranded hairpin loops, phenol-chloroform extracted and ethanol precipitated. The purified DNA was amplified with library primers (SL2 and AL2) and purified on an agarose gel; DNA of ~200 bp was cut out and extracted from the gel matrix. The purified DNA was PCR amplified again to add a T7 promoter to one end using T7SL2 and AL2 primers. The complexity of the library was estimated to be ~10⁹, therefore, the entire genome was represented by multiple overlapping fragments in both orientations (i.e. “plus” and “minus” strands).

Primer sequences

SLLT

5' -X-CGTAGTGAGGCACGTCTGTTTTTCAGACGTGCCTCACTACGT-3'
X=5' Phosphate

ALLT

5' -X-GCAATGCGTCAAGCTCAGTTTTCTGAGCTTGACGCATTGCT-3'
X=5' Phosphate

SL2

“Sense” PCR oligo:
5' -GGGCAGACGTGCCTCACTAC-3'

T7SL2

“Sense” PCR oligo with T7 promoter:
5' -GATCTGTAATACGACTCACTATAGGGCAGACGTGCCTCACTAC-3'

BIOT7

5' -XGATCTGTAATACGACTCACTATAGGG-3'
X=5' Biotin

AL2

“Antisense” PCR oligo (with 5’ phosphate)
5' -XCTGAGCTTGACGCATTG-3'
X=5' Phosphate

T7ALSP2

5' -GCGTCAAGCTCAGGATCTGTAATACGACTCACTATAGGG-3'

Selection protocol.

To make a single-stranded circular version of the library, the DNA was amplified using a biotinylated forward primer (BIOT7) and 5'-phosphorylated reverse primer (AL2), immobilized on Streptavidin beads (BioRad), denatured with NaOH (0.15 M, 2 min, 22 °C) and eluted. The eluted single-stranded DNA was neutralized with 3M sodium acetate, pH 5.2, ethanol precipitated and subsequently annealed to a splint DNA oligonucleotide (T7ALSP2) and incubated with T4 DNA ligase to circularize the library. The ligated product was then incubated with Taq DNA polymerase and dNTPs at 72 °C for 30 minutes to make a double-stranded circular library (Fig. S1B).

The circular double-stranded DNA library was transcribed using T7 RNA polymerase to make rolling-circle transcripts (Fig. S1B). Long RNA transcripts, corresponding to multimers of the library were purified on denaturing PAGE, eluted and precipitated. The RNA was resuspended in TE buffer (10 mM Tris.Cl, pH 8.0, 1 mM EDTA), heated at 94 °C for 3 minutes and allowed to cool to room temperature over 10 minutes. To begin the selection step, the RNA was incubated in 140 mM KCl, 10 mM NaCl, 5 mM MgCl₂, 50 mM Tris.Cl, pH 7.4, 10 mM DTT, ~1 unit RNasin (Promega) for 1 hour at 37 °C. The RNA was then resolved by denaturing PAGE and RNA of ~400 nts was excised and eluted. The isolated RNA was reverse-transcribed and PCR amplified using primers BIOT7 and AL2 to restart the selection cycle.

In pilot experiments carried out with a hammerhead ribozyme motif, we observed generation of rolling circle transcripts and their self-cleavage into unit-length RNAs, as well as the expected dimer, trimer, tetramer, and larger multimers, as had been described earlier (S1). A mutated inactive form of the hammerhead ribozyme produced only long concatemers. A simplified pilot experiment with these two hammerhead constructs resulted in a 50-fold enrichment of active sequences relative to inactive sequences in a single round of selection, demonstrating that this approach could be successfully applied to the selection of novel ribozymes.

Selected sequences.

Cloning and sequencing of the round 12 library revealed 11 distinct sequences (Fig. S2). Nearly half of the clones corresponded to the same sequence, which maps to an olfactory receptor gene *OR4K15*, but is antisense with respect to the coding strand. We assayed a representative clone from each of the 11 sequence groups individually in unit-length forms. The olfactory receptor gene sequence was shown to be active and three other sequences showed robust self-cleavage activity: a sequence complementary to an *IGF1R* intron, a portion of a LINE1 sequence, and a sequence mapping to the second intron of the *CPEB3* gene (in sense orientation) (Fig. S3). We also sequenced a number of clones from rounds 10 and 11. Examples of the four active sequences identified from round 12 were present in these rounds, however a number of sequences with lower activity were also recovered. For instance, a G-U rich sequence with a predicted secondary structure resembling miRNA precursors and an α-satellite centromeric repeat were among the notable clones. Except for the *CPEB3* clone, we did not test self-cleaving activity of these clones without their flanking primer sequences or having reverted any mutations that had accumulated during the course of the selection.

CPEB3 ribozyme notes.

We measured the rate of self-cleavage of the CPEB3 ribozyme isolated from a number of species. The mouse sequence, which differs from the human sequence at only two of 78 positions (Fig. 1F), self-cleaves somewhat faster than the human sequence ($k_{obs}=2.9 \pm 0.6 \text{ hr}^{-1}$). Elephant, rabbit and opossum sequences also self-cleave several times faster than the human ribozyme. In contrast, the ribozyme derived from the rat sequence self-cleaves somewhat more slowly ($0.164 \pm 0.005 \text{ hr}^{-1}$).

We decided to test whether the CPEB3 and HDV ribozymes fold into the same secondary structure because of their biochemical similarities: *i*) the CPEB3 and HDV ribozymes have similar metal dependencies (S2); *ii*) like the HDV ribozymes (S3, S4), the CPEB3 ribozyme has a relatively flat pH profile and similar solvent kinetic

isotope effect; *iii*) in both cases most of the sequence required for self-cleavage is downstream of the cleavage site and *iv*) like the HDV ribozymes, the CPEB3 ribozyme has a stable fold because it can function in 4M urea at moderate magnesium concentration (Fig. S6).

The human gene has a single nucleotide polymorphism (SNP) at position 36 (U36C). This SNP occurs in our unselected library with a 50:50 ratio to the wild-type (WT); however, almost all of the isolated CPEB3 clones (40/41) had the U36C mutation, indicating that C was positively selected during the course of the selection (Fig. S5). We measured the *in vitro* activity of the C36 variant relative to the WT and observed a ~2.6-fold higher self-cleavage rate, suggesting a functional relevance for this SNP. The bovine and the rat forms of the ribozyme form an A1-U36 base-pair at this position. Introducing the G1A mutation within the context of the human ribozyme slightly increases the rate of self-cleavage (Fig. S7), while mutating the G1-U36 wobble pair to an A1-C36 wobble pair slows the ribozyme cleavage rate by about 3-fold. Taken together, the CPEB3 ribozyme self-cleaves faster when the first nucleotide downstream of the cleavage site forms a Watson-Crick base-pair with position 36.

In order to test the possibility that general nuclear factors could lead to a constitutively higher level of self-cleavage activity *in vivo*, we incubated the ribozyme with transcription-competent HeLa nuclear extract (Promega). Both the WT and the U36C variant ribozymes self-cleaved in the presence of the extract, however the rate of cleavage was not significantly altered.

We used RNABOB (<http://selab.wustl.edu/cgi-bin/selab.pl?mode=software#rnabob>) to search the four CPEB genes and all conserved mammalian sequences (PhastCons sequences flanked by 20 bp upstream and downstream genomic sequences) (S5) for sequences that could fold into the HDV-like fold. Other than the CPEB3 ribozyme, no sequences that could fold into the same secondary structure were identified. Similarly, we searched the clones from an *in vitro* selection of self-cleaving ribozymes based on a random library (S6). None of the clones could be threaded through the HDV secondary structure without violating secondary structure constraints. These results suggest that the CPEB3/HDV fold is conserved only in the *CPEB3* intron and that it is unlikely that it could arise multiple times independently.

Cleavage assays.

All kinetics experiments were started by addition of Mg²⁺ to ³²P body-labeled RNA and aliquots were stored in excess EDTA and 7 M urea until analysis. Products of self-cleavage were resolved in denaturing polyacrylamide gels and analyzed using a GE Healthacre (Amersham Bioscience) phosphorimager and ImageQuant software.

5' RACE.

To detect the presence of the cleaved version of CPEB3 ribozyme in different tissues, 5' RACE (SMART-RACE kit, Clontech) was performed on total RNA samples isolated from mouse organs (a gift from Dr. B. Seed's laboratory) and human testis and placenta RNA (Ambion). Mouse-specific RT primer AL142 5'-GAGACGAGGAAAATGT was annealed to 200 ng (less than 200 ng for hippocampus) of total RNA in 2.5 μL water, in the absence of any buffer. An additional 2.5 μL, containing the rest of the reaction components were added to the mixture on ice, then the reactions were incubated at 42 °C for 1.5 hrs. The reaction mixture contained SMART IIA oligo 5'-AAGCAGTGGTATCACGCAGAGTACGCGGG that hybridized to the 3' end of the cDNA (5' end of RNA) during the reverse transcription. PowerScript reverse transcriptase adds several untemplated C residues to the 3' end of cDNA, and the template switching mechanism (SMART) enables reverse transcription from the SMART IIA oligo which adds its complement to the 3' end of cDNA. Forty to 44 cycles of PCR were performed on the samples using universal ribozyme-specific primer AL141 5'-TACAGCAGAATTCTCGCAGATT, which was used for both mouse and human RNA samples, and nested

universal primer NUP 5'-AAGCAGTGGTATCAACGCAGAGT, which is a truncated SMART IIA oligo sequence. The RT primer used with human samples was AL122_C 5'-AATTGCTGTTGACTTGGG.

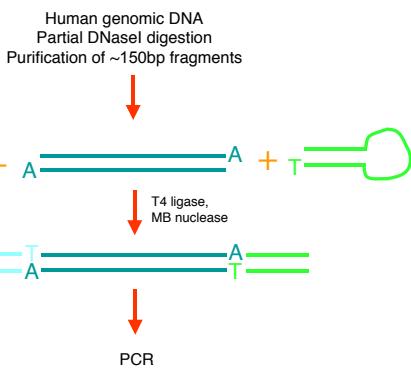
The expected PCR product of 100 bp was observed in human brain and placenta as well as mouse brain, cortex, testis and spleen; however, we did not detect the 100 bp band in mouse liver, hippocampus and cerebellum. PCR was also performed with the original RNA without including reverse transcriptase to assess the non-specific background; the 100 bp band corresponding to the RACE product was not detected. Following electrophoresis, the relevant band was eluted from the gel, cloned using the TOPO TA cloning system (Invitrogen) and sequenced. Sequences from human testis and placenta, mouse total brain, cortex, testis and spleen confirmed the presence of the cleaved CPEB3 ribozyme with the cleavage site corresponding to the in vitro determined location (Fig. S8, Table S1).

Detection of ribozyme cleavage by RT-PCR in human RNA.

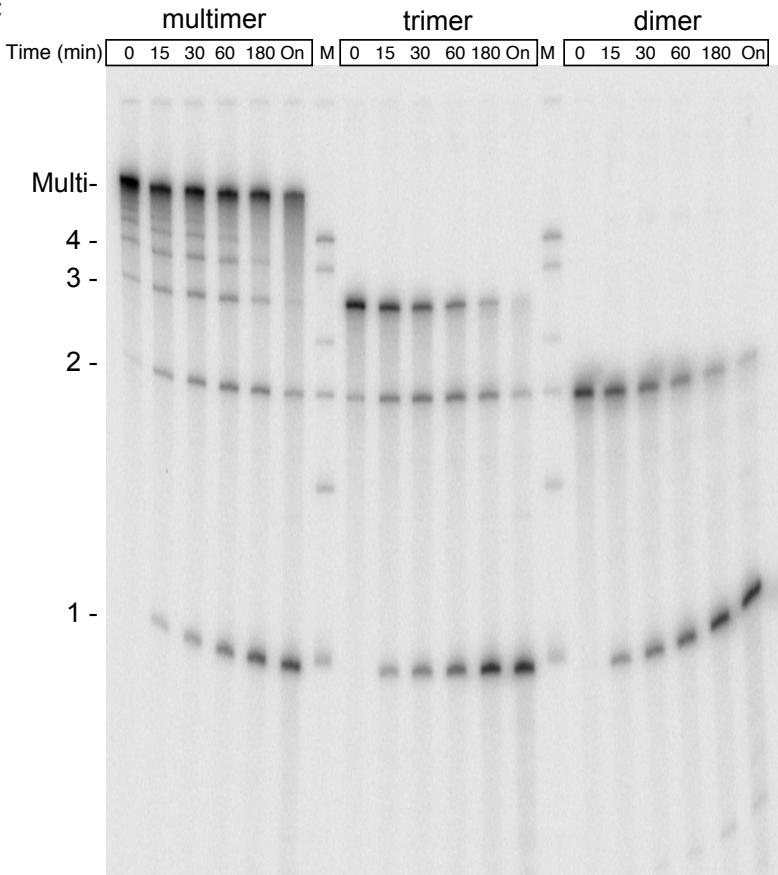
We used two primer sets to amplify the CPEB3 ribozyme from human RNA and measure the ratio of cleaved to uncleaved ribozyme. The first set of primers (AL121 and AL122C) amplifies a segment spanning the cleavage site of the ribozyme; therefore it can only amplify the uncleaved ribozyme. The second primer set (AL120B and AL122C) does not span the cleavage site, therefore can amplify both the cleaved and uncleaved ribozyme. We carried out PCR to detect the CPEB3 ribozyme using these primer pairs on reverse-transcribed human RNAs. Analysis of the results showed differences in the ratio of cleaved to uncleaved ribozyme in different RNA samples, with placenta having the lowest and brain the highest (Fig. S9).

Fig. S1

A



C



B

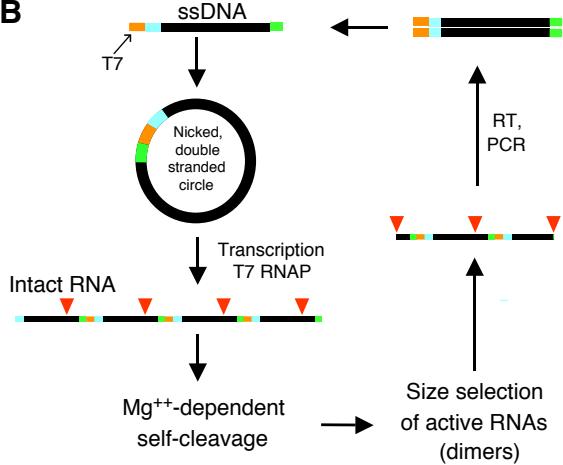


Figure S1. Library construction, *in vitro* selection scheme and pool activity. **(A)** Construction of the library from human genomic DNA. Partially digested human genomic DNA was size selected, incubated with Taq DNA polymerase to add 3'-terminal adenosines and subsequently ligated to double-stranded hairpin primers using T4 DNA ligase. The single-stranded loops of the ligated primers were digested using mung bean nuclease (MB) to produce a linear double-stranded DNA pool that was subsequently amplified by PCR to add a T7 promoter sequence and purified by agarose gel electrophoresis. **(B)** In vitro selection scheme. A single-stranded DNA pool with a 5' T7 RNA polymerase promoter (T7) was circularized and converted into nicked double-stranded transcription template. Transcription yielded concatemeric RNA, potentially encoding sites of self-cleavage (red triangles). After Mg²⁺-induced cleavage, RNAs of dimer length were isolated, reverse-transcribed (RT) and amplified (PCR) to start the next round of selection. **(C)** Mg²⁺-dependent self-cleavage of round 12 pool. Length of the starting material (in terms of library copies in transcripts) is indicated at the top of the figure. Incubation times and product sizes are indicated above and to the left of the PAGE gel, respectively (M, marker; ON, over-night).

Fig. S2.

OR4K15

1	GTTAACGGTC ATAGGCCATG GAAACTAGGA GCACCATTG CGTGCCAGTG
51	AAGAAATGAA CAAAGAAAAT CTGGGCCAGG CAGGCATCAA AAGAAATAGT
101	CTTGCCTCA ACCAGAAAGT CTGTAATCAT TTTAAGGGTA

CPEB3

1	TACGTGCTGC TTAGTTCTTA AATTGAGGA TCAAGGGGAT AACAGGGGGC
51	CACAGTAGAA GCGTTCACGT CGCAGCCCC GTCAGATTCT AGTGAATCTG
101	CGAATTCTGC TGCATATCTC CTTACCCAAC CACCTAANTC AAACA

IGF1R

1	GTTGGCTTT ACAGCTAGAT GCACTCTCCA CAAGCCAACA GGAGTCGCGA
51	GTGAGAGCCC TCGAACACAT CCAGTGTGAA CATGCATTAA GCAAACATTC
101	CTGAGCCCTG TTTTGTGCCA GGCTGTTCTC AGTGCTCAA

Rc1

1	GTGGAATGAG AAAAATACAA GACAGTCTAA GATCTAACCT TTGTCGCCT
51	TTCTAGGCTT ATGTCTTAAT ACTCCCGAAG TGGCACAAAGC TAGGTTATTT
101	CTAGTTCTTA ACAGACTACT GTGCTTTGC ACAATGCTCT A

Rb12

1	GTTACACATG TTTTCCGTGT AACATGTGCA TGTGGTATTG CCGTATGCAA
51	GTATTCCGTG TAAAGTTGT ACACAAAGTG GTCATACGCA TTCATTCCGT
101	GTAATGTCTG CATGCGATAT TGTATACGCA ATGCGTTCCCT A

Rd2

1	TAAGANGTGA ATATCTACTC CGTTGGTTA TTCTCCATCG GAGAGACATT
51	CTTCAGATGC ACTTACTCAG TCATCTCGAA TCCAGAGAAC AGAATATGGA
101	CCTAGAAACA TAACATGAAA TTATGTTGT GAGTATCTGA CGTA

Rb9

1	TACGTTAAGG ACCAAAGTTA TGCCATGCAT ATATTGCTTG GCACACATCA
51	AGTACTCAAC AGCTGACAGC AGCTGTTACC TTGACGATCA TCACTGGTAA
101	AAGCCGATGA TGTGTTGTG GAAGAAAGCC ACAGATCACA GTAGCT

Rb10

1	TACGTCATAT ACGTGTGTGC ATGCATGTGC ATGTATGTGT GCACGTGTGT
51	GTGCATGTGT GTGTATGTGT ATATACTAAC CTGTGTGCAT TGTGTGACTA
101	TGTGCGCATG CATGTGCATG TGTGTGCATA TACGTG

Re9

1	GTAGAGACAG CAGTAGTAGC CCCACGAGGG TGCTCGGCC TTTGCCATAC
51	AGCTAACTCA GTACATAACG CAGCTCTCA GTACAGGCTT GATGACAGTT
101	GGTAATTAAT TGCATCACCG TAACCTCTGT A

Line1

1	TACGTAGGTG GTTTCCCT GATAGTGCTA AGGGGGCCTG GAAGTTGGAA
51	CTGGCAGAG CTCAACACAG CGTGGCAAAG TGGCTGTGGC TAGACTGCCT
101	CTCTGGATTC CTCGTCACTC AGCAGGGCAT CTCTGAGAGA AATGCC

Figure S2. Round 12 clones. Cloned sequences of the selected self-cleaving genomic ribozymes without primer sequences and including any mutations that were accumulated during the selection.

Fig. S3.

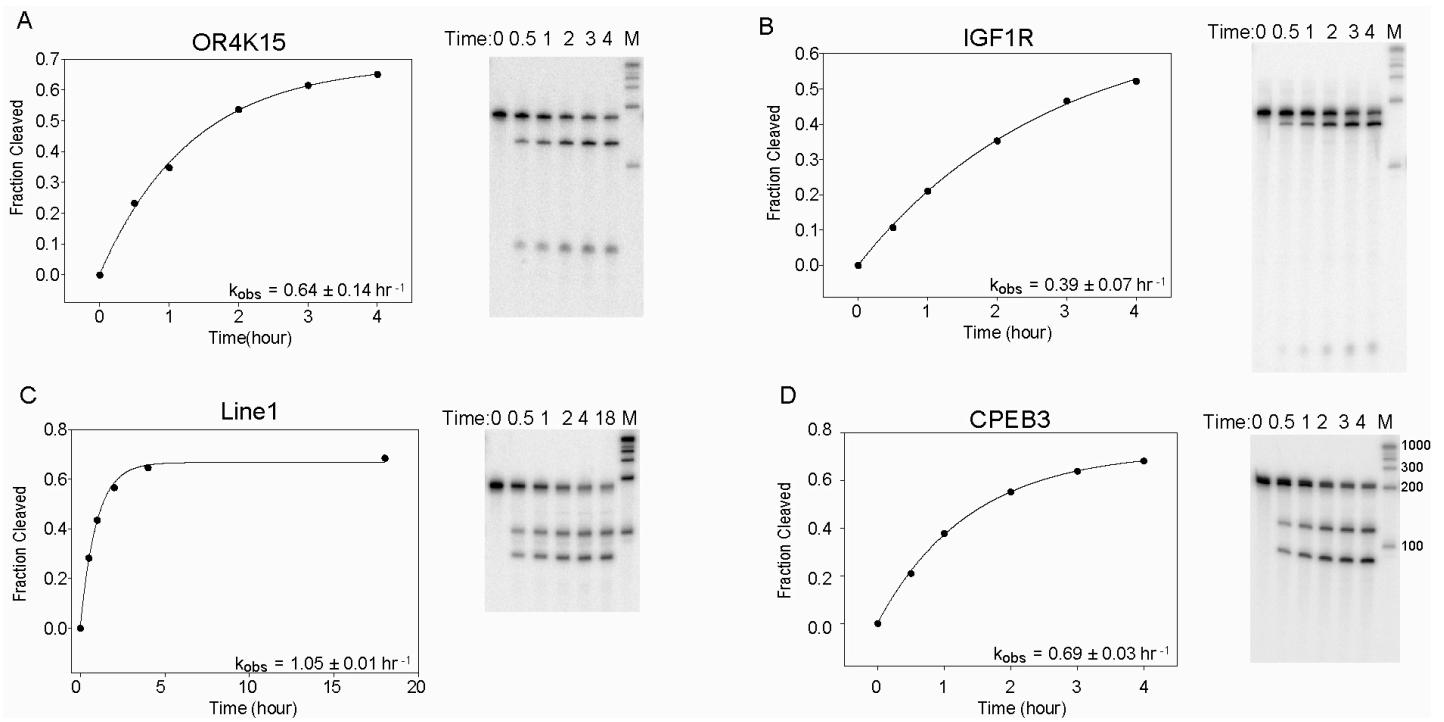


Figure S3. Self-cleavage of the most active clones from the selection. Self-cleavage of sequences that match the genomic sequences of (A) olfactory receptor *OR4K15*, (B) *IGF1R*, (C) Line 1 retroposon and (D) *CPEB3* genes. Representative gels (right) and time courses (left) are shown for each sequence (Fig. S2). First order rate constants were averaged over three experiments and are shown as inserts.

Fig. S4.

	
CPEB3	GGCTGTTGTT TTATTGCTTT TATTGCTATA GGATGATTTG CATAAAAGCT GCTGCTTAGT ub2	----- GGGTTT TATTGCT--A GGATGATTTG CATAAAAGCT GCTGCTTAGT
ra8	----- ----- ----- ----- ----- ----- ----- ----- -----	
ph12	----- ----- ----- ----- ----- ----- ----- ----- -----	
tc9	----- ----- ----- ----- ----- ----- ----- ----- -----	
pd3	----- ----- ----- ----- ----- ----- ----- ----- -----	
rb7	----- ----- ----- ----- ----- ----- ----- ----- -----	
re6	----- ----- ----- ----- ----- ----- ----- ----- -----	
pf10	----- ----- ----- ----- ----- ----- ----- ----- -----	
od6	----- ----- ----- ----- ----- ----- ----- ----- -----	
og9	----- ----- ----- ----- ----- ----- ----- ----- -----	
nd5	----- ----- ----- ----- ----- ----- ----- ----- -----	
rc8	----- ----- ----- ----- ----- ----- ----- ----- -----	
	
	65 75 85 95 105 115	
CPEB3	TCCTAAATTC AAGGATCAAG GGGATAAACAG GGGGCCACAG CAGAACCGTT CACGTCGCAG ub2	CCCTAGATTC AAGGATCAAG GGGATAAACAG GGGGCCACAG CAGAACCGTT CACGTCGCAG
ra8	TCCCTAAATTC GAGGATCAAG GGGATAAACAG GGGGCCACAG TAGAACCGTT CACGTCGCAG	
ph12	TCCCTAAATTC AAGGATCAAG GGGATAAACAG GGGGCCACAG TAGAACCGTT CACGTCGCAG	
tc9	TCCCTAAATTC AAGGATCAAG GGGATAAACAG GGGGCCACAG TAGAACCGTT CACGTCGCAG	
pd3	TCCCTAAATTC AAGGATCAAA GGGATAACTG GGGGCCACGG CAGAACCGTT CACGTCGCAG	
rb7	TCCCTAAATTC AAGGATCAAG GGGATAAACAG GGGGCCACGG TAGAACCGTT CACGTCGCAG	
re6	TCCCTAAATTC AAGGATCAAG GGGATAAACAG GGGGCCACAG TAGAACCGTT CACGTCGCAG	
pf10	TCCCTAAATTC AAGGATCAAG GGGATAAACAG GGGGCCACAG CAGAACCGTT CACGTCGCAG	
od6	TCCCTAAATTC AAGGATCAAG GGGATAATAG GGGGCCACAG TAGAACCGTT CACGTCGCAG	
og9	TCCCTAAATTC AAGGATCAAG GGGATAAACAG GGGGCCACGG CAGAACCGTT CACGTCGCAG	
nd5	TCCCTAAATTC AAGGATCAAG GGGATAAACAG GGGGCCACAG CAGAACCGTT CACGTCGCAG	
rc8	TCCCTAAATTC AAGGATCAAG GGGATAAACAG GGGGCCACAG TAGAACCGTT CACGTCGCAG	
	
	125 135 145 155 165 175	
CPEB3	CCCCCTGTCAG ATTCTGGTGA ATCTGCGAAT TCTGCTGTAT ATCTCCTTAC CCAACCACCC ub2	CCCCCTGTCAG ATTCTGGTGA ATCTGCGAAT TCTGCTGTAT ATCTCCTTAC CCAACCACCC
ra8	CCCCCGTCAG ATTCTAGTGA ATCTGCGAAT TCTGCTGCTAT ATCTCCTTAC CCAACCACCT	
ph12	CCCCCGTCAG ATTCTAGTGA ATCTGCGAAT TCTGCTGCTAT ATCTCCTTAC CCAACCACCT	
tc9	CCCCCGTCAG ATTCTAGTGA ATCTGCGAAT TCTGCTGTAT ATCTCCTTAC CCAACCACCC	
pd3	CCCCCGTCAG GTTCTAGTGA ATCTGCGAAT TCTGCTGTAT ATCTCCTTAC CCAACCACCC	
rb7	CCCCCGTCAG ATTCTAGTGA ATCTGCGAAT TCTGCTGTAT ATCTCCTTAC CCAACCACCT	
re6	CCCCCGTCAG ATTCTAGTGA ATCTGCGAAT TCTGCTGTAT ATCTCCTTAC CCAACCACCC	
pf10	CCCCCGTCAG GTTCTAGTGA ATCTGCGAAT TCTGCTGTAT ATCTCCTTAC CCAACCACCC	
od6	CCCCCGTCAG ATTCTAGTGA ATCTGCGAAT TCTGCTGTAT ATCTCCTTAC CCAACCACCC	
og9	CCCCCGTCAG ATTCTAGTGA ATCTGCGAAT TCTGCTGTAT ATCTCCTTAC CCAACCACCC	
nd5	CCCCCGTCAG ATTCTAGTGA ATCTGCGAAT TCTGCTGTAC ATCTCCTTAC CCAATTACCC	
rc8	CCCCCGTCAG ATTCTAGTGA ATCTGCGAAT TCTGCTGTAT ATCTCCTTAC CCAACCACCC	
	
	185 195 205	
CPEB3	AAGTCAAACA GCAATTGAAA TGAACAGCT ub2	----- ----- ----- ----- ----- ----- ----- ----- -----
ra8	AANTCAAACA ----- ----- ----- ----- ----- ----- ----- -----	
ph12	----- ----- ----- ----- ----- ----- ----- ----- -----	
tc9	AAGTCAAACA ----- ----- ----- ----- ----- ----- ----- -----	
pd3	AAGTCAAACA ----- ----- ----- ----- ----- ----- ----- -----	
rb7	AAGTCAAACA ----- ----- ----- ----- ----- ----- ----- -----	
re6	----- ----- ----- ----- ----- ----- ----- ----- -----	
pf10	AAGTCAAAGCA ----- ----- ----- ----- ----- ----- ----- -----	
od6	AANTCAAAC- ----- ----- ----- ----- ----- ----- ----- -----	
og9	AAGTCAAACA ----- ----- ----- ----- ----- ----- ----- -----	
nd5	AAGTCAAACA ----- ----- ----- ----- ----- ----- ----- -----	
rc8	AANTCAAACA ----- ----- ----- ----- ----- ----- ----- -----	

Figure S4. Alignment of the 12 independent clones of the CPEB3 ribozyme. Independent isolates of the CPEB3 ribozyme are aligned with the genomic sequence (CPEB3). The cleavage site is located between position 89 and 90. The human SNP is at position 125. (-) marks gaps in the isolated sequences, relative to the genomic reference. The minimum sequence required for self-cleavage is underlined.

Fig. S5.

	5	15	25	35	45	55
CPEB3	GGCTGTTGTT	TTATTGCTTT	TATTGCTATA	GGATGATTTG	CATAAAAGCT	GCTGCTTAGT
ub2		G.G...				
ra8					T.CG.	
uf12					T.CG.	
ph10					T.CG.	
od12					T.CGA	C..
na7					T.CA.	
ne3					T.CG.	
rc12					T.CG.	
nf1					CG.	T..
pc10					CA.	
pa2					CA.	
ra10					CG.	
ng2					C..	
od5					CG.	
ng1					CG.	
ob4					G.	
oa5					G.	
ph12					G.	
ph6					G.	
pf12					GC	T..A
uh5					G.	
uh9					G.	
ya3					G.	
ng3					G.	
yal1					G.	
yc1					G.	
ng12					A.	
tc9						
pd3						
nh9						
oc12						
rb7						
re6						
pf10						T..
pb2						
oc7						
od6						
og9						
oh3						
nd5						
rc8						

	65	75	85	95	105	115
CPEB3	TCCTAAATTC	AAGGATCAAG	GGGATAACAG	GGGGCACAG	CAGAACGTT	CACGTCGCAG
ub2	C.....	G.....
ra8	G.....	T.....
uf12
ph10	..T.	T.....
od12	G.....
na7	T.....
ne3	G.....
rc12	T.....
nfl	T.....
pc10	T.....
pa2
ra10	G.....	T.....
ng2	T.....	T.....
od5	T.....
ng1	T.....
ob4	T.....
oa5	T.....
ph12	T.....
ph6	G.....	G.....
pf12
uh5	T.....
uh9	T.....
ya3	-.....	T.....
ng3	T.....	A.....
ya11	A.G.....

yc1G.....TG.....
ng12
tc9T.....T.....
pd3A.....T.....G.....
nh9
oc12T.....
rb7G.....T.....
re6T.....
pf10
pb2
oc7A.....
od6T.....T.....
og9G.....
oh3T.....
nd5G.....
rc8T.....

	125	135	145	155	165	175
CPEB3	CCCTGTCAG	ATTCTGGTGA	ATCTGCGAAT	TCTGCTGTAT	ATCTCCTTAC	CCAACCACCC
ub2C.
ra8C.AC	T
uf12C.G.....A
ph10A	T
od12CAT
na7CA
ne3CG.....A	T.....T
rc12CAA
nflCATT
pc10CAT
pa2CA
ra10CAC	T
ng2CAC
od5CA.....GT
ng1CA
ob4CAC
oa5CG.....A	T
ph12CAA
ph6CAAT
pf12CG.....A
uh5CA.....A	T
uh9CA.....A	T
ya3CA	TT
ng3CG.....AT	T
ya11CG.....A
yc1CA
ng12CAT	T
tc9CA
pd3CG.....A
nh9CAT	T
oc12CG.....AT	T
rb7CA	TT
re6CA
pf10CG.....A
pb2CA
oc7CG.....A	T
od6CT.....A
og9CA
oh3CG.....A
nd5CACT	TT
rc8CA	TT

		185		195		205
CPEB3	AAGTC	AAACA	GCAATTGAAA	TGAACAGCT		
ub2	-----					
ra8	..N.....					
uf12	..N.....					
ph10					
od12					
na7					
ne3T....					
rc12	C.....G...					
nf1					
pc10					

pa2-----
ra10-----
ng2-----
od5T.....-----
ng1-----
ob4-----
oa5-----
ph12	-----
ph6T.....-----
pf12T.-----
uh5-----
uh9-----
ya3-----
ng3G-----
ya11-----
yc1	..N.....-----
ng12	...TGG...-----
tc9	-----
pd3-----
nh9	...TGG...-----
oc12G-----
rb7-----
re6	-----
pf10G..-----
pb2-----
oc7T.-----
od6	..N.....-----
og9-----
oh3	T.....-----
nd5-----
rc8	..N.....-----

Figure S5. Alignment and conservation of all 41 cloned CPEB3 sequences. All clones of the CPEB3 ribozyme were aligned and the positions at which they differ from the genomic reference sequence (CPEB3) are indicated by the identity of the mutation; (.) marks positions at which the clones match the reference sequence and (-) marks gaps in the isolated sequences, relative to the genomic reference. The minimum sequence required for self-cleavage is underlined. The numbering is the same as in Fig. S3.

Fig. S6.

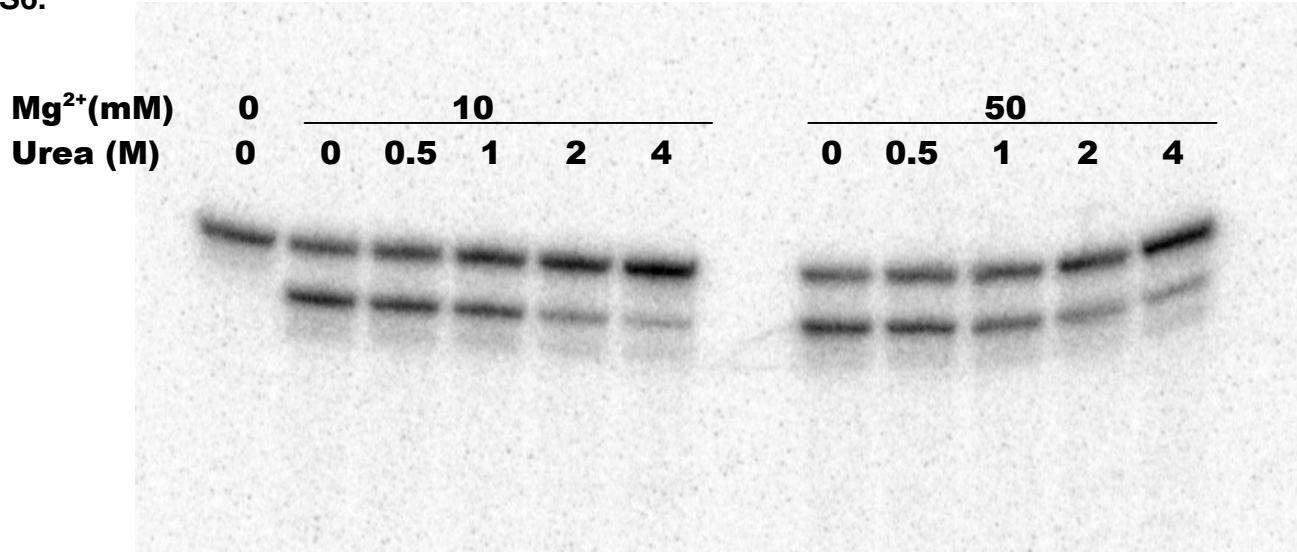


Figure S6. Self-cleavage of the CPEB3 ribozyme in urea. All samples were incubated for 2 hours at 37 °C in 10 (left) or 50 mM Mg²⁺ (right). The concentration of urea used is indicated above each lane.

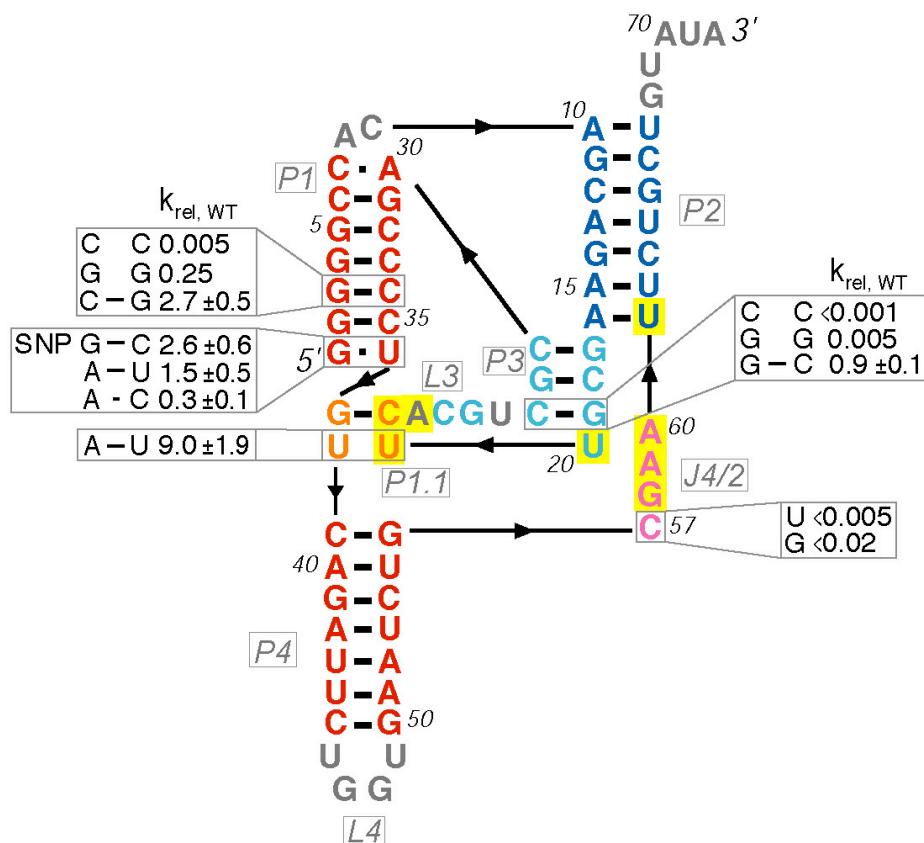
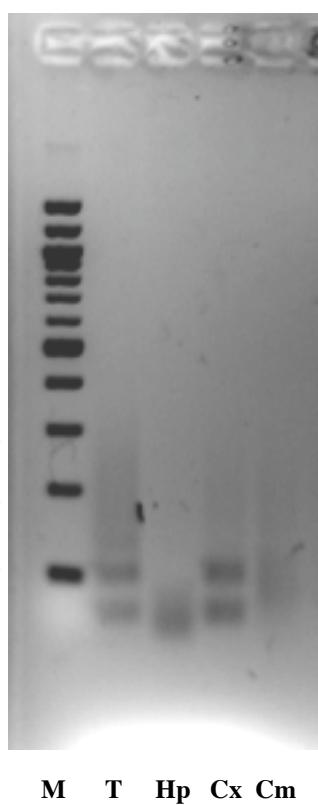


Figure S7. Mutant CPEB3 ribozymes and their rates of self-cleavage, relative to the wild-type sequence. 5' marks the cleavage site. Yellow boxes indicate phosphate positions sensitive to phosphorothioate substitution.

Fig. 8.

A



Mouse RACE product:

AAGCAGTGGTATCAACGCAGAGTACGCCGG*GGGGGCCACAGCAGAAGCGTTACGTCGCGGCCCCTGTCAGATTCTGGCG
AATCTGCGAATTCTGCTGT

Human RACE product:

AAGCAGTGGTATCAACGCAGAGTACGCCGG*GGGGGCCACAGCAGAAGCGTTACGTCGCAGCCCCTGTCAGATTCTGGTG
AATCTGCGAATTCTGCTGT

Figure S8. Sequence and results of SMART-RACE experiments. Mouse tissues were taken into ice-cold RNAlater solution (Ambion). Whole brain was dipped into RNAlater for an hour, dissected, then stored in RNAlater. The following organs and tissues were collected: whole brain, cortex, cerebellum, hippocampus, testis, liver, spleen. The tissue samples (20-30 mg) were disrupted in buffer RLT (Qiagen) by passing through a syringe and then homogenized using Qiashredder column (Qiagen). Total RNA was isolated by RNeasy plus kit (Qiagen) that uses a dedicated column for removal of genomic DNA. The RNA isolation reagents contained EDTA and no DNase treatment was performed to avoid the possibility of ribozyme self-cleavage during these steps. **(A)** An example of SMART-RACE performed with mouse total RNA samples from different tissues: M-100 bp marker; T-testis, Hp- hippocampus, Cx- cortex, Cm- cerebellum. The expected length of the product is 100 bp. No 100-bp band is observed in -RT control reactions. **(B)** Sequences of mouse and human 100-bp RACE products; nucleotides in red font denote CPEB3 ribozyme; green font denotes SMART IIA oligo, asterisk indicates the ribozyme cleavage site. Positions where the human and mouse genomic sequences differ are underlined.

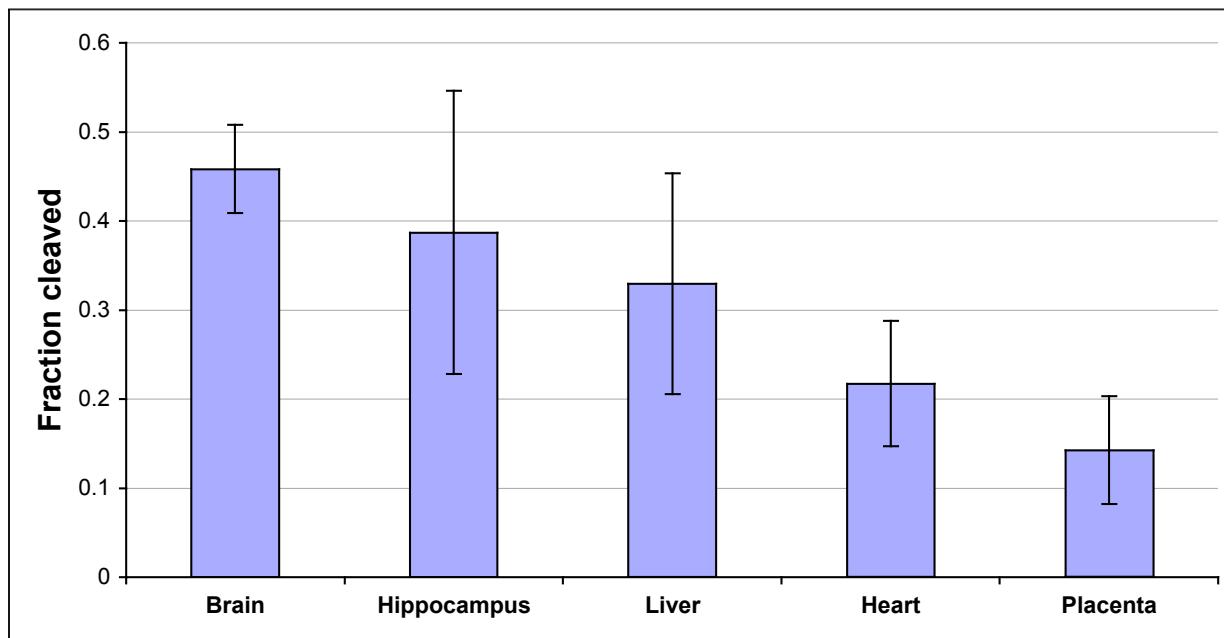


Figure S9. RT-PCR analysis of the human CPEB3 ribozyme. Human RNA isolated from brain, hippocampus (a gift from Dr. David Gerber, Galenea Corp.), liver, heart (left ventricle), and placenta were reverse transcribed using random hexamers as primer. Approximately 1.0 µg RNA was mixed with random hexamers (50 µM final reaction concentration) in water, heated briefly to 72 °C, cooled to ambient temperature, then reverse transcribed by addition of Superscript III (Invitrogen) using the manufacturer's recommended reaction components in 20 µl reactions. The reactions were sequentially incubated at ambient temperature, 37 °C, 42 °C, and 55 °C. Following heat inactivation, 1 µl of reverse-transcribed samples was PCR amplified (in 50 µl reactions) for 31-35 cycles using the primer pair AL121 + AL122C to detect uncleaved ribozyme, or AL120B + AL122C to detect both cleaved and uncleaved ribozymes. The graph represents analysis of three or four experiments. Error bars represent standard errors. All RNAs except hippocampus were purchased from Ambion. The sequence of primers is as follows: AL121 5'
GTTCCCTAAATTCAAGGATCAAGGGGATAA, AL122C 5' AATTGCTGTTGACTTGGG, AL120B 5'
CCACAGCAGAAGCGTTCACGT.

Table S1. Sequences from the tissue-specific 5' RACE experiments.

Organism	Tissue	Total clones sequenced	Clones showing cleaved ribozyme
Mouse	brain	9	6
Mouse	brain cortex	10	8
Mouse	testis	7	6
Mouse	spleen	8	7
Human	testis	12	11
Human	placenta	6	6

References

- S1. S. L. Daubendiek, E. T. Kool, *Nat Biotechnol* 15, 273 (1997).
- S2. S. R. Das, J. A. Piccirilli, *Nat Chem Biol* 1, 45 (2005).
- S3. S. Nakano, D. M. Chadalavada, P. C. Bevilacqua, *Science* 287, 1493 (2000).
- S4. I. Shih, M. D. Been, *Proc Natl Acad Sci U S A* 98, 1489 (2001).
- S5. A. Siepel et al., *Genome Res* 15, 1034 (2005).
- S6. K. Salehi-Ashtiani, J. W. Szostak, *Nature* 414, 82 (2001).